Shining a Light Inside the Black Box

Part 3 of 4

1

Speakers

- Fred Ducca, FHWA
- Bill Davidson, PB Americas
- Bill Woodford, AECOM Consult

2

Series Schedule

- Four sessions of two hours each
  - "Motivations & Data": February 12th, 2008 at 2:30 PM EST
  - "Model Testing": March 11th, 2008 at 2:30 PM EST
  - "Transportation Supply & Travel Distribution": April 8th, 2008 at 2:30 PM EDT
  - "Translating Results Into Insights for Decision Makers": May 13th, 2008 at 2:30 PM EDT
- Please submit questions to chat pod to Dave Schmitt
- Questions will be answered at the end of each session

3

Speaker

- Bill Davidson, PB Americas

4

Transportation Supply

"I knew I was going to take the wrong train, so I left early"

- Yogi Berra

The Importance of Transportation Supply

- The representation of transportation supply via networks, coding conventions and path-building impact the model's ability to grasp traveler behavior
- Many model problems can be traced to the representation of the transportation system though the network/zone "grain", link characteristics and paths
- The representation of supply should reflect the actual operation of the transportation system

7

Model Problems Traced to Supply

Problem	Potential Supply Cause(s)
Severe over-/under-assignment in localized area Severely low speeds	Excessively large zone "grain" Misrepresentation of speed or capacity
Over-assignment on major roadways and under-assignment on minor roadways	Excessively high freeway speeds or low arterial speeds Illogical volume-delay functions Insufficiently defined path cost function
Insufficient closure of equilibrium assignment in future years	High growth in very large zones

8

Model Problems Traced to Supply

Problem	Potential Supply Cause(s)
High positive bias constants needed to calibrate drive-access transit trips	Access connectors incorrectly defined Incorrect access connector, auto or transit speeds
High negative bias constants needed to calibrate walk-access transit trips	Access connector or transit speeds too fast (manually coded?) Auto speeds too slow Over-estimation of transit coverage

9

Transportation Supply as Represented in Travel Models

Aspect	Represented via…
Physical facilities and services	Highway and transit networks (including speeds, capacities and coding conventions)
Access to facilities and services	Zone "grain" Centroid connectors Access connectors
How people traverse through system	Paths (including path cost criteria and volume-delay functions), network connectivity

10

Link Speed and Capacity

- In addition to being functions of facility types, speed and capacity are functions of the physical characteristics of the link
- Speed and capacity should be related directly to these characteristics, instead of using facility type/area type associations

11

Link Speeds

- For free-flow conditions, the posted speed limit may work in many situations, other ideas include the "pull" speed on freeways, signal data and "progression" speeds on arterials
- For congested conditions, data from car probes or traffic management centers should work for most situations

12

Determining Link Speeds:
Southeast Florida Regional Model

- This model uses an automated procedure to compute free-flow speeds based on:
  - Posted speed for uninterrupted flow
  - Posted speed, signal spacing & timing data for urban interrupted flow
- The data requirements include posted speed for all links, location of signalized intersections, distance between signals for signal controlled areas, green to cycle ratio, cycle length and assessment of progression to estimate delay adjustment factor
- The procedure is based on NCHRP Report 387 *Planning Techniques to Estimate Speeds and Service Volumes for Planning Applications* (1997)

13

Determining Link Speeds:
North Central Texas Council of Gov'ts (Dallas)

- Free-flow speed is based on the speed limit and the mid- or end-link traffic control devices
- The delay from traffic controls is a function of the mid-/end-link control, control type, functional class and area type

14

Link Capacity

- Capacity is a function of many physical attributes, including:
  - Number of lanes and lane widths,
  - On-street parking and presence of medians,
  - Intersection and curb/access controls,
  - Turn lanes and terrain and others…
- It should be plausibly related to these attributes instead of facility type/area type associations

15

Determining Link Capacity
Ohio Department of Transportation

- The automated routine computes a saturation flow based on the physical characteristics of the link with HCM2000 as the data source
- There are two levels of detail, one with basic detail and another with extra detail
  - Level 1 includes curb-to-curb width, facility type, area type and number of lanes
  - Level 2 includes all Level 1 data plus:
    - The number of intersection turn and through lanes
    - Existence of a mid-link median turn lane
    - Existence of on-street parking
    - Intersection control and terrain type

16

Transit Networks

- In addition to good route coding, these elements of transit supply should be compared against actual data
  - Mixed-flow running times should be related to auto speeds in an explainable way
  - Fixed-guideway running times should be consistent with actual operating times
  - Walk access should be represented as ubiquitous access when possible
  - Drive access should be represented by connectors to parking facilities
  - Transfer areas or stations should utilize coding conventions allow reasonable connections while disallowing impossible ones

17

Relating Transit Speeds to Auto Speeds

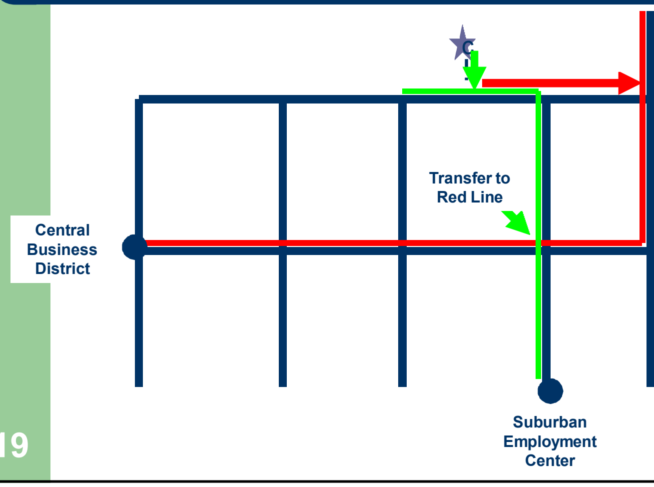
- Transit speeds are slower than auto speeds because transit vehicles have to stop to pick-up and discharge passengers, so:

  TransitTime = AutoTime + B*X
  where B is the number of stops and
  X is the average dwell time per stop

- Other relationships such as: TransitSpeed = (AutoSpeed * 2) fail to relate to auto speeds in an explainable way

18

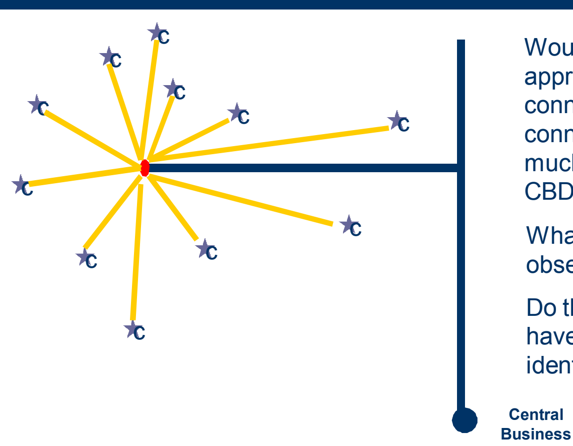
A Reason for Ubiquitous Walk-Access



Ubiquitous walk-access allows longer walk to avoid transferring to Red Line

19

Drive-Access



Would backtracking logic be appropriate to remove connectors that allow a connection to transit that is much further away from the CBD than the origin?

What is the maximum observed drive-access time?

Do the access connectors have some type of easily identifiable shape?

20

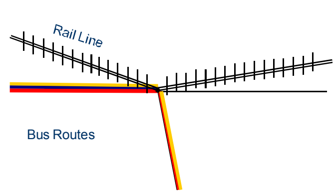
Highway and Transit Network Connectivity

- Networks should represent connectivity at a level sufficiently high to reflect meaningful differences between alternatives
- For highway networks, this might mean:
  - Micro-coding ramp configurations or lane configurations to avoid over-stating queue or weaving conditions

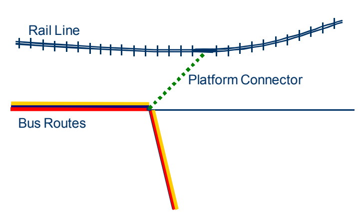
21

Highway and Transit Network Connectivity (cont'd)

- For transit networks, this might mean:
  - Reflecting the correct frequency and route
  - Ensuring that major transfer areas have the correct stop locations
  - Micro-coding transit platforms to avoid over-stating transit connectivity (under-stating travel time)



No Micro-coding



Micro-coded Station

22

Path Testing

- Paths are reliant on good networks and reasonable cost criteria
- An easy test is to review paths from key origins to major destinations, such as major residential areas to employment centers
- This test validates network coding, speeds and path-building assumptions in addition to the routing
- The transit path-builder needs to be consistent with the mode choice model in structure and the weights of the travel time components

23

Testing the Representation of Transportation Supply

This test…	Examines the Reasonableness of…
Assignment of survey trip tables	Network coding conventions Network connectivity Speeds
Path-checking	Network connectivity Path cost function Speeds

24

Network/Zone "Grain"

- Network should consist of all facility types of at least one level below the desired assignment fidelity
  - The zones are the "holes" in this roadway system
- Zonal boundaries should be balanced with available data, but not subservient to it
  - Examples: large census block groups in downtown areas and rural areas in model edges

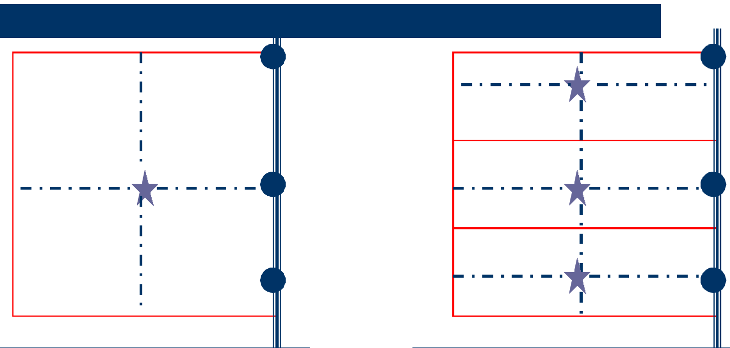
25

The Zonal "Future Year Problem"

- We expect growth in large vacant or rural zones near the model edge; these typically have low levels of activity in the base year
- Large activity levels in the future may overwhelm the simplified representation of supply in these areas causing undesired effects in forecasting
- Consequently, the zonal area should account for expected growth rather than the low activity level in the base year

26

The Zonal "Future Year Problem"
Example



This large zone has experienced large growth. A corridor study is looking to address the travel congestion, but the large zone hampers the model's ability to reflect station activity and transit accessibility

In this graphic, the large zone has been subdivided to better capture the accessibility and station access. The same example can occurs in highway studies when large zones has access to multiple freeway interchanges.

27

So What Does This All Mean?

- Realistic representation of the key characteristics of transport supply (travel time, cost, capacity, connectivity) is vital for developing models that describe effectively travel behavior.
- Moreover, one of the key purposes for travel forecasting models is to evaluate the adequacy of regional transportation infrastructure. This application is ***meaningless*** without an accurate depiction of the operation of those resources.
- Modeled supply must be continuously checked:
  - Coded data (e.g., number of lanes, bus headway, connectivity).
  - Functional relationships (capacity as a function of physical features, bus time as a function of highway time and boardings)
  - Paths and skim matrices

28

Speaker

- Bill Woodford, AECOM Consult

29

Travel Distribution

30

"If you don't know where you're going, you'll wind up somewhere else"

- Yogi Berra

31

Why is Distribution Important?

- The major input to mode choice and assignment models
- Errors in distribution become "correction factors" in mode choice and assignment. Some examples:
  - Huge bias constants and decision rules in mode choice
  - Excessive county-level factoring disguised as "turning penalties" or "bridge penalties"
  - 45 mph free-flow freeways and 70 mph arterials

32

Impact of Distribution Errors on Demand

Estimated Demand/Travel Patterns

	CBD	Urban	Suburbs	Tech Center	Rural	Total
CBD	1,000	1,000	-	-	-	**2,000**
Urban	40,000	1,000	-	1,000	-	**42,000**
Suburbs	7,000	1,000	10,000	35,000	2,000	**55,000**
Tech Center	1,000	3,000	3,000	1,000	-	**8,000**
Rural	1,000	19,000	7,000	3,000	-	**30,000**
Total	**50,000**	**25,000**	**20,000**	**40,000**	**2,000**	**137,000**

Observed Demand/Travel Patterns

	CBD	Urban	Suburbs	Tech Center	Rural	Total
CBD	1,000	-	-	1,000	-	**2,000**
Urban	7,000	10,000	21,000	3,000	1,000	**42,000**
Suburbs	35,000	1,000	5,000	12,000	2,000	**55,000**
Tech Center	2,000	-	1,000	4,000	1,000	**8,000**
Rural	5,000	-	-	20,000	5,000	**30,000**
Total	**50,000**	**11,000**	**27,000**	**40,000**	**9,000**	**137,000**

Circumferential freeway to Tech Center overloaded by 300%

Inbound freeway to Tech Center underloaded by 80%

Subway trips over by 400%

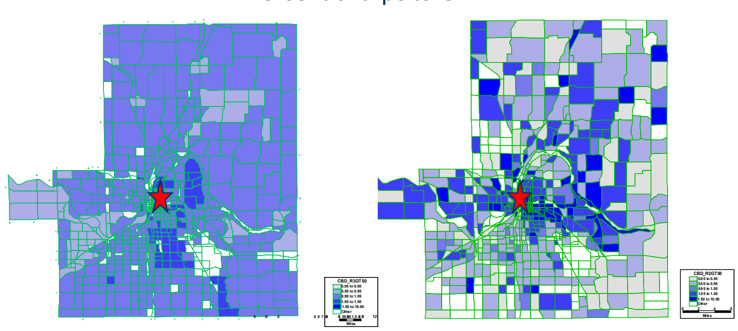
Commuter rail under by 80%

Radial freeway loads up too near CBD

33

Distribution presents an idealized view of travel patterns

Percent of trips to CBD



Model

CTPP

34

Common Problems with Distribution Models

- Over-reliance on transportation as the driver of distribution
- Insufficient disaggregation of trip purposes
- Adverse impact of production/attraction (P/A) balancing
- The use of aggregate calibration/validation measures
- The lack of sufficient data for testing

35

The Over-Reliance on Transportation

- Almost all models use transportation-related attributes as key determinants
- Distribution is a fairly complex social behavior with longer-term impacts
  - Available & location desirable housing
  - Employment location based on supply, availability, cost & external decisions

36

Distribution:
The Real-Life & Model Disparity

Real-Life

- Home/work interaction does not change with each transportation change
- Long-term choice
- What people do I want to live nearby? Where is my job located?

Model

- Home/work interaction always changes with each transportation change
- Intra-day travel
- Which job am I going to today?

37

Key Issues in Distribution

- Design:
  - Gravity vs. destination choice design
  - Degree of aggregation of purposes and socio-economic classes
  - Single- vs. double-constrained trip end balancing
  - Highway vs. highway + transit impedances
- Validation approach
  - Trip length frequency
  - Trip patterns at ***appropriate*** level of disaggregation

38

Gravity vs. Destination Choice

- Logit Destination Choice is a generalized form of the gravity model
  - Negative exponential friction factors in gravity model
  - LN of attractions used as size variable in destination choice

$$T_{ij} = P_{ij} \times \frac{\exp(-c_{ij}t_{ij})A_j K_{ij}}{\sum \exp(-c_{ij}t_{ij})A_j K_{ij}} = P_{ij} \times \frac{\exp(-c_{ij}t_{ij} + \ln(A_j) + \ln(K_{ij}))}{\sum \exp(-c_{ij}t_{ij} + \ln(A_j) + \ln(K_{ij}))}$$

- Best approach depends on software application and whether generalized form is required.
- K factors may be required for non-transportation effects

39

Specific Characteristics of Land Use/ Population/ Employment are Important

- Worker characteristics by residential and work place location are unevenly distributed across the region: ***i.e., Occupation and Income***
- Shopping travel patterns varies by type of retail
  - Neighborhood grocery store—short trip lengths
  - Regional malls—medium trip lengths
  - Destination retail—IKEA, Nordstrom's—long trip lengths
- Other travel patterns
  - Social-Recreation—Visit neighbors, visit family, recreation-gym work out vs. recreation “Travel” soccer all different
  - Airports—draw from entire region

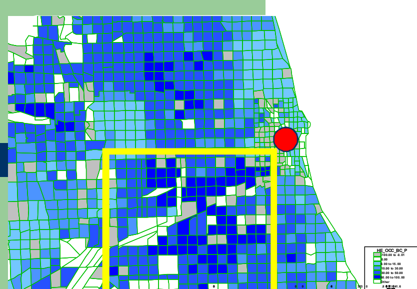
40

Specific Characteristics of Travel May Explain Discontinuities

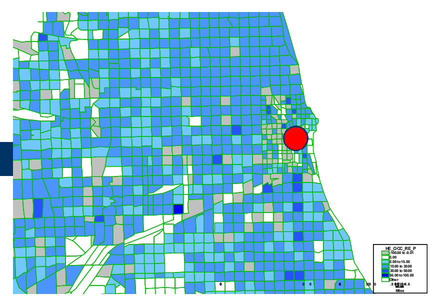
- CTPP data can be used to study the spatial relationship between home and workplace by socio-economic status
  - Part 1 – Residence tables
  - Part 2 – Place-of-work tables
  - Occupation tabulations are in Table 003 (both Parts 1 & 2)
- The following thematic maps display this information by showing the proportion of workers in occupation group by zone

41

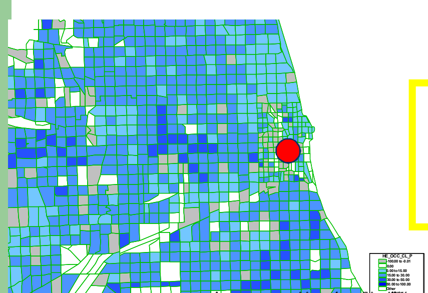
Chicago Workers by Occupation Class – Home-End



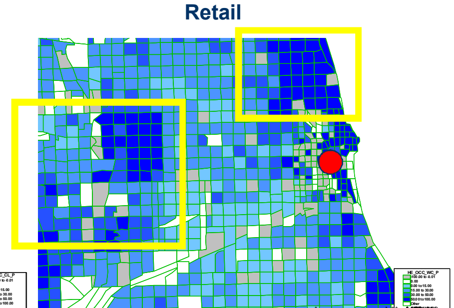
Blue Collar



Retail



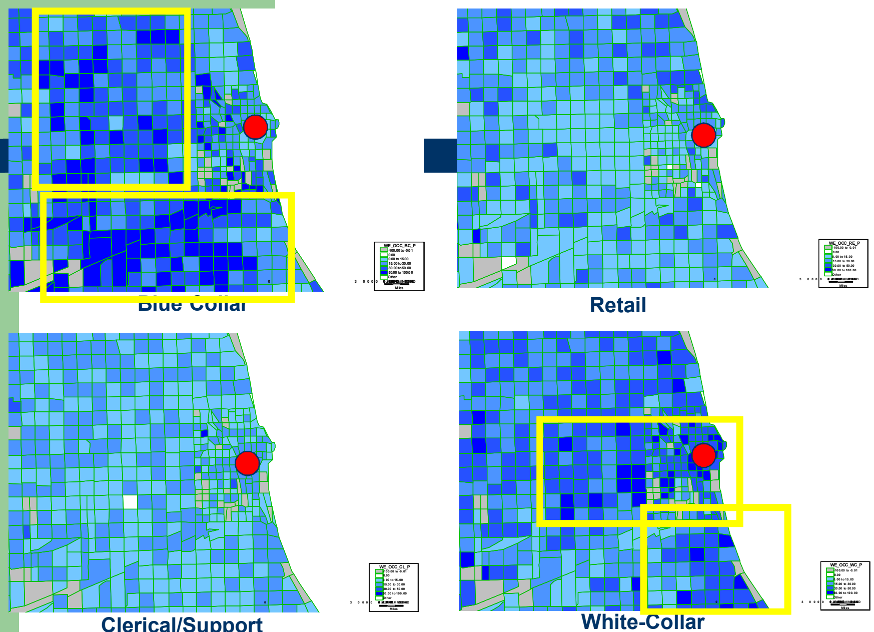
Clerical/Support



White-Collar

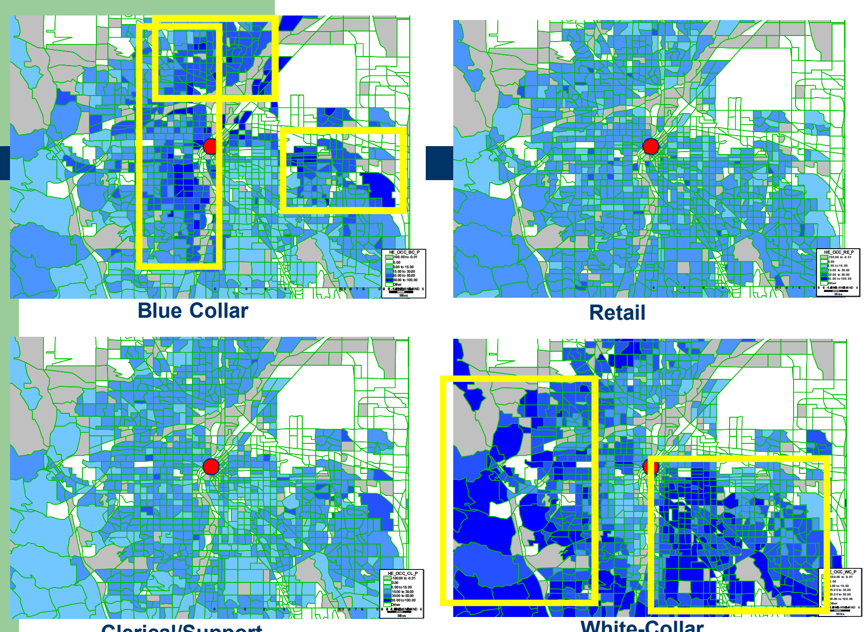
42

Chicago Workers by Occupation Class – Work-End



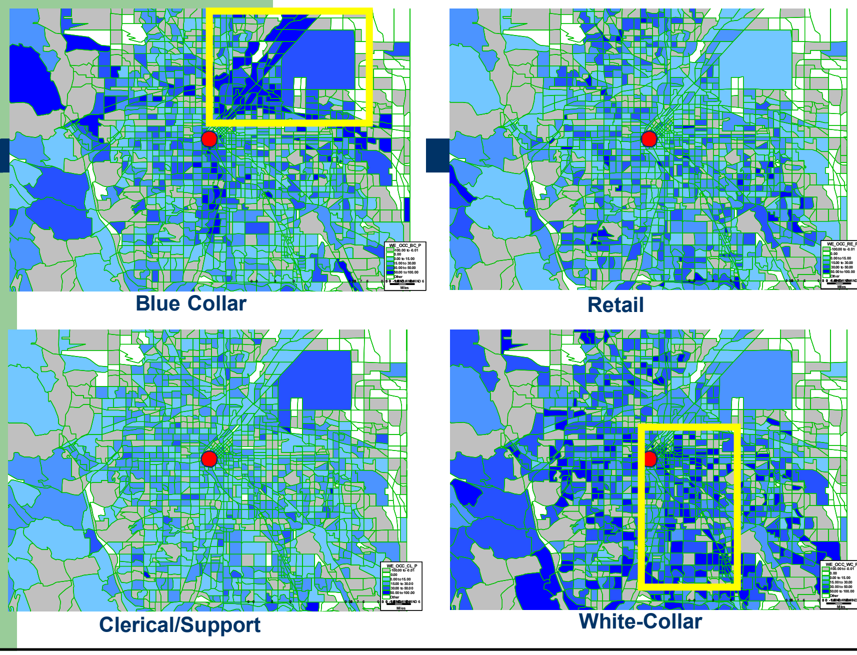
43

Denver Workers by Occupation Class – Home-End



44

Denver Workers by Occupation Class – Work-End



45

Decision to Use Single- or Double-Constrained Trip End Balancing

- Dependent on nature of input SE Forecasts:
  - Do worker and workplace forecasts explicitly consider transportation?

  Some forecasts of employment imply substantial transportation flows that may or may not occur.

  - Are shopping trips really related to market forecasts?

  Successful shopping centers or shopping malls attract higher numbers of customers per square foot. This is a function of local demographics, retail offerings, and competition. Is this factored into the forecasting process?

46

Effects of Socio-Economic Data on Doubly-Constrained Distribution

Workers/Jobs by County (Year 2000, Projected in 2003)			
County	Workers	Jobs	Difference
San Francisco	444,851	634,430	-189,579
San Mateo	403,083	395,890	7,193
Santa Clara	959,071	1,092,330	-133,259
Alameda	697,882	751,680	-53,798
Contra Costa	483,898	361,110	122,788
Solano	179,517	123,210	56,307
Napa	67,111	66,840	271
Sonoma	229,307	205,220	24,087
Marin	140,955	122,960	17,995
Region	3,605,675	3,753,670	-147,995

Non-zero implies ***minimum*** level of cross county travel with doubly constrained model. Negative value = Net importer of workers / Positive value= Net exporter of workers.

47

Source: SCVTA

Are these forecasts usable for double-constrained models?

Difference between Year 2000 Estimates of Workers and Jobs		
County	Projection '03	Census 2000
San Francisco	-189,579	-204,967
San Mateo	7,193	-16,865
Santa Clara	-133,259	-180,698
Alameda	-53,798	-40,603
Contra Costa	122,788	90,682
Solano	56,307	46,224
Napa	271	-6,474
Sonoma	24,087	13,579
Marin	17,995	-2,221
Region	-147,995	-301,343

2003 projection of 2000 conditions understated worker/job differences reported after release of 2000 CTPP

48

Source: SCVTA

Are these forecasts usable for double-constrained models?

Difference between Workers and Jobs (2030)			
County	Projection '03	Projection '05	Projection '07
San Francisco	-268,180	-270,390	-300,760
San Mateo	-35,900	-42,490	-44,120
Santa Clara	-168,270	-253,670	-40,650
Alameda	-24,170	-56,770	1,070
Contra Costa	168,330	123,940	110,870
Solano	106,170	51,890	84,900
Napa	-5,990	1,780	-14,010
Sonoma	-11,920	18,390	-49,970
Marin	2,120	5,520	-13,510
Region	-237,810	-421,800	-266,180

Substantial variation between different projection series.

Source: SCVTA

49

EXAMPLE OF P/A BALANCING IMPACT

Before Balancing

Base		CBD	Tech Center	Growth Suburb	Static Suburb	Total	Target	Target Grw
	CBD	75	5	50	60	190	190	0.0%
	Tech Center	200	400	100	200	900	900	0.0%
	Growth Suburb	20,000	7,500	80,000	50,000	157,500	157,500	0.0%
	Static Suburb	60,000	2,000	100,000	120,000	282,000	282,000	0.0%
	Total	80,275	9,905	180,150	170,260	440,590	440,590	
	Target	81,078	17,334	257,049	85,130	440,590		
	Target Growth	1.0%	75.0%	42.7%	-50.0%	0.0%		

After Balancing

Columns		CBD	Tech Center	Growth Suburb	Static Suburb	Total
	CBD	76	10	74	30	190
	Tech Center	150	569	109	75	902
	Growth Suburb	17,959	12,778	104,451	22,365	157,553
	Static Suburb	62,893	3,978	152,415	62,660	281,945
	Total	81,078	17,334	257,049	85,130	440,590

Change

Columns		CBD	Tech Center	Growth Suburb	Static Suburb	Total
	CBD	2%	93%	48%	-49%	0%
	Tech Center	-25%	42%	9%	-63%	0%
	Growth Suburb	-10%	70%	31%	-55%	0%
	Static Suburb	5%	99%	52%	-48%	0%
	Total	1%	75%	43%	-50%	0%

50

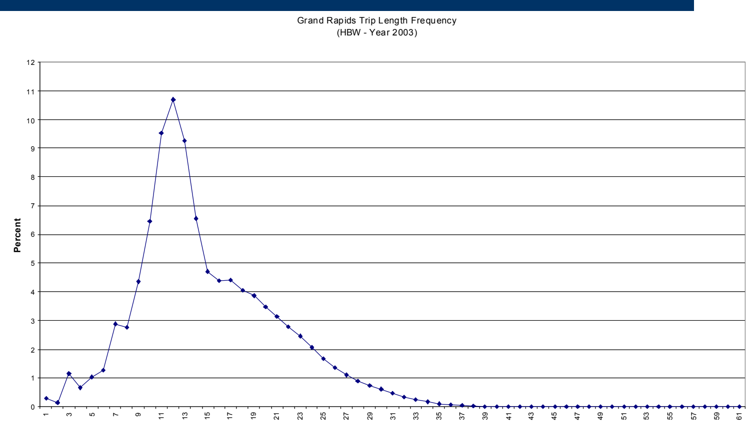
Calibration/Validation Measures

- Calibrating distribution models using average trip lengths by purpose does not ***alone*** provide sufficient information about the travel markets produced by the distribution model
- Detailed inspection of the person trip flows is required

51

Good TLFs but Poor Travel Patterns

Example: TLFs



52

Good TLFs but Poor Travel Patterns

Example: Modeled Travel Patterns

Model	Ottawa	South Central	Airport / Mall	Near SE	Steelcase	Near SW	West Kent	CBD	Others	Total
Ottawa	-	-	-	-	-	-	-	-	-	**-**
GVSU	-	-	-	-	-	-	-	-	-	**-**
Far SW	1,549	1,344	913	1,478	568	1,703	1,294	1,271	4,290	**14,409**
Medical	480	443	115	492	103	498	335	193	1,366	**4,025**
South Central	2,742	2,754	2,363	2,968	1,349	3,138	2,646	3,094	8,691	**29,746**
Far SE	448	476	511	480	250	513	463	596	1,762	**5,499**
Airport/Mall	2,187	2,363	898	2,889	654	2,380	1,856	1,316	7,900	**22,444**
Near SE	3,077	2,968	2,889	3,552	1,603	3,665	3,284	3,915	10,259	**35,212**
Steelcase	1,314	1,349	654	1,603	418	1,456	1,154	972	4,274	**13,195**
Near SW	3,808	3,138	2,380	3,665	1,456	4,218	3,400	3,351	10,551	**35,968**
West Kent	3,167	2,646	1,856	3,284	1,154	3,400	3,617	2,898	10,459	**32,482**
CBD	3,044	3,094	1,316	3,915	972	3,351	2,898	2,109	10,814	**31,513**
East Central	762	791	728	948	372	890	773	861	3,036	**9,160**
Amway	226	244	163	300	89	255	224	207	1,009	**2,718**
Near NE	1,314	1,309	1,484	1,483	726	1,517	1,693	1,938	6,246	**17,710**
Plainfield Heights	1,038	973	997	1,275	528	1,226	1,234	1,395	4,393	**13,059**
Uptown	923	867	888	1,089	493	1,096	1,127	1,305	3,469	**11,256**
Near NW	964	823	632	1,048	382	1,037	1,185	984	4,113	**11,168**
Far NE	794	747	783	840	381	862	1,025	1,037	3,767	**10,236**
Far NW	668	509	[illegible]	638	318	735	911	[illegible]	2,746	**7,992**
Total	**28,506**	**26,840**	**20,155**	**31,948**	**11,814**	**31,940**	**29,119**	**28,324**	**99,147**	**307,792**

Good TLFs but Poor Travel Patterns

Example: Observed Travel Patterns

CTTP scaled to Model	Ottawa	South Central	Airport / Mall	Near SE	Steelcase	Near SW	West Kent	CBD	Others	Total
Ottawa	-	-	-	-	-	-	-	-	-	**-**
GVSU	-	-	-	-	-	-	-	-	-	**-**
Far SW	516	1,634	1,438	1,031	1,621	2,250	670	1,512	3,641	**14,311**
Medical	9	144	96	60	147	72	58	61	148	**795**
South Central	433	7,196	7,061	4,343	3,985	2,769	1,317	4,020	6,331	**37,456**
Far SE	39	989	1,275	697	605	319	246	556	2,395	**7,123**
Airport/Mall	44	692	1,915	1,274	487	419	286	986	1,620	**7,723**
Near SE	290	2,466	7,574	9,811	2,951	2,894	2,165	7,699	7,463	**43,313**
Steelcase	143	845	1,330	1,190	1,203	863	382	937	1,142	**8,036**
Near SW	1,507	2,589	4,614	4,345	3,640	8,644	2,316	3,627	6,951	**38,233**
West Kent	898	1,096	2,780	3,489	1,999	2,391	7,465	5,344	6,458	**31,919**
CBD	180	691	2,419	2,370	711	1,073	1,016	3,493	2,728	**14,681**
East Central	96	574	1,694	1,372	429	372	364	1,566	3,212	**9,680**
Amway	6	112	211	203	78	102	72	286	614	**1,684**
Near NE	93	1,011	2,611	2,208	1,221	1,037	1,679	3,109	10,476	**23,444**
Plainfield Heights	157	647	1,875	1,886	603	764	1,278	2,135	6,749	**16,093**
Uptown	188	506	1,859	2,532	794	1,042	1,570	3,003	4,797	**16,289**
Near NW	93	404	1,109	1,008	537	966	2,405	1,738	4,870	**13,130**
Far NE	60	395	1,155	965	516	572	1,146	1,377	5,660	**11,846**
Far NW	131	224	[illegible]	1,027	599	732	1,330	[illegible]	6,090	**12,037**
Total	**4,880**	**22,218**	**41,746**	**39,811**	**22,127**	**27,280**	**25,766**	**42,621**	**81,344**	**307,792**

Checking Distribution Patterns

- If meaningful, estimated distribution should be compared to CTPP 2000 at a level as disaggregate as possible
  - District-to-district being the most aggregate level (districts are sub-county for most urban areas)
- Very important for key travel markets
  - Suburban → CBD
  - Suburban → Suburban
  - Corridor movements

55

The Lack of Sufficient Data

- Household surveys with very small sample sizes are not sufficient to be used in a meaningful verification of estimated travel patterns, as cells would be represented by five or fewer observations
- Better and more data is needed to be able to make meaningful comparison

56

Possible Solutions

- Be aware what models aren't capturing (can't capture?) in travel distribution
- Include aspects that can be captured
- Learn from land-use models that account for real estate & employer decisions
- Explore alternative strategies of capturing non-transportation elements

57

Other Remedies for Problematic Distribution Models

- Short-term
  - Use data entirely (CTPP 2000, Montreal/Toronto)
  - Embark on more complete disaggregation of distribution models
  - Apply K-factors as needed, but beware
- Long-term
  - Research into alternative techniques such as Urban Sim

58

What Should You Be Doing?

- Set up a separate task in your UPWP to identify a data collection plan that relates to the model development and testing
- Base your model's representation of supply on physical characteristics and data
- Validate your representation of supply, especially speeds and paths
- Validate your distribution model using detailed comparisons of travel patterns
- Explore alternative strategies of capturing non-transportation elements in your distribution model

59

Questions

60

Special Thanks

- To those that contributed examples, insights and experiences:
  - Santa Clara Valley Transportation Authority
  - Association of Bay Area Governments
  - Interurban Transit Partnership
  - Florida Department of Transportation, District 4
  - North Central Texas Council of Governments
  - Ohio Department of Transportation
  - Chicago Metropolitan Agency for Planning
  - Denver Regional Council of Governments

61

References

- *CAP2000: Capacity Calculator Program Documentation*. Ohio DOT. July 2001.
- *Dallas-Fort Worth Regional Travel Model (DFWRTM): Model Description*. North Central Texas Council of Governments Transportation Department. July 2007.
- *Southeast Florida Regional Planning Model VI: 2000 and 2030 Models – Technical Report 2 (DRAFT) – Model Calibration & Validation*. The Corradino Group. May 2007.
- *SPD2000: Network Speed Calculator Documentation*. Ohio DOT. July 2001.

62